

# **THIRD-PARTY CALL CONTROL TYPE SIMULTANEOUS INTERPRETATION SYSTEM AND METHOD THEREOF**

## **BACKGROUND OF THE INVENTION**

5

This application claims the priority of Korean Patent Application No. 10-2002-0068580 filed on November 6, 2002, in the Korean Intellectual Property Office, the disclosure of which is incorporated herein by reference.

### 10 1. Field of the Invention

The present invention relates to a third-party call control type simultaneous interpretation system and method, and more particularly, to a system and method capable of providing interactive simultaneous interpretation services to talkers and listeners connected with the system through wired/wireless communication networks.

### 15 2. Description of the Prior Art

As international exchange has continued to expand, opportunities to converse with or talk on the telephone to foreigners who use another language have increased. Thus, an interpretation system for performing smooth communication with foreigners is now required.

As an interpretation system used to communicate with foreigners, Korean Patent  
20 Laid-Open Publication No. 2002-0030693 (entitled "Voice interpretation service method and voice interpretation server") discloses a method wherein the voice of a user is first transmitted to a voice interpretation server and a translated voice is then returned to the user through a telephone capable of using a mobile internet access service, as shown in FIG. 1.

In such a case, the voice interpretation method has an advantage in that an  
25 interpretation service can be provided conveniently through the voice interpretation server regardless of the time and position if the user utilizes a predetermined terminal. However, there are problems in that the user should hire or purchase the terminal for the interpretation service from a provider and the method is not suitable to a means for communicating with foreigners who are remotely located because it is a one-way interpretation service between the  
30 user and the voice interpretation server.

In order to solve these problems, Korean Patent Laid-Open Publication No. 2002-54192 (entitled "System and method for automatically interpreting telephone information for foreigners") discloses a system of automatically interpreting telephone information, as an interactive interpretation system for performing communication with foreigners who are remotely located and use a different language. The system is configured in such a manner that when a foreigner user asks a question in his/her own language, the question is automatically interpreted and then is transmitted to a native operator and the response of the native operator to the question is then automatically interpreted and transmitted to the foreigner user.

However, when the foreign user connects with the simultaneous interpretation system through a wired/wireless telephone, the system for automatically interpreting telephone information is configured to connect the call of the foreign user to the native operator connected with the simultaneous interpretation system. Thus, the system can substantially provide the interpretation services only to the foreign user and the native operator. Therefore, there is a limitation in that the simultaneous interpretation system is not suitable to an interpretation means for communicating between any two users, who use different languages, (e.g., a Korean user A and an English user B) with each other.

## SUMMARY OF THE INVENTION

The present invention is conceived to solve the aforementioned problems. An object of the present invention is to provide a simultaneous interpretation system and method for allowing users, who use different languages and are remotely located, to conveniently communicate with one another.

According to an aspect of the present invention for achieving the object, there is provide a third-party call control type simultaneous interpretation system, which comprises a CTI board for establishing a traffic channel between a talker and a listener, a CTI control module for generating an event in response to a button signal input through the CTI board to control the CTI board as a job unit capable of performing a basic telephone action, an interpretation module for recognizing a voice of the talker/listener input through the CTI board and translating the voice into a predetermined language, and a main control module for

controlling an action of the CTI control module in accordance with a predetermined interpretation scenario.

According to another aspect of the present invention, there is provided A third-party call control type simultaneous interpretation method, which comprises a telephone connection step of establishing a traffic channel between a talker and a listener when the talker connects with a simultaneous interpretation system; an automatic interpretation step of, when an event is generated in a CTI control module in response to a button signal input by the talker or listener through a CTI board, translating an input voice of the talker or listener into a predetermined language in response to the generated event based on a predetermined interpretation scenario; and an interpretation transmission step of controlling the CTI board in accordance with the interpretation scenario and transmitting the translated voice to the other party in accordance with the interpretation scenario.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The above and other objects and features of the present invention will become apparent from the following description of preferred embodiments given in conjunction with the accompanying drawings, in which:

FIG. 1 is a view showing a configuration of a conventional simultaneous interpretation system;

FIG. 2 is a view illustrating a conventional simultaneous interpretation method;

FIG. 3 is a view schematically showing a configuration of a network for use in a third-party call control type simultaneous interpretation system according to the present invention;

FIG. 4 is a view schematically showing a configuration of the third-party call control type simultaneous interpretation system according to the present invention;

FIG. 5 is a view illustrating operations of a working section shown in FIG. 4;

FIG. 6 is a view showing an example of an interpretation scenario according to the present invention; and

FIG. 7 is a flowchart illustrating an entire process of the third-party call control type simultaneous interpretation method according to the present invention.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hereinafter, the configuration and operation of a third-party call control type simultaneous interpretation system and method according to the present invention will be explained in detail with reference to the accompanying drawings.

FIG. 3 is a view schematically showing a configuration of a network for use in the third-party call control type simultaneous interpretation system according to the present invention. Referring to FIG. 3, when a talker 100 connects with a third-party call control type simultaneous interpretation system 500 through a public switched telephone network 700 (hereinafter, referred to as "PSTN") and a private automatic branch exchange 900 (hereinafter, referred to as "PBX"), the simultaneous interpretation system 500 receives a telephone number of a listener 300 from the talker 100 to establish the predetermined traffic channel. Then, the system automatically translates the voice of the talker 100 input through the established traffic channel and transmits the translated voice of the talker to the listener 300, and also automatically translates the voice of the listener 300 and transmits the translated voice to the talker 100.

For example, a case where a traffic channel is established between a Korean talker 100 and an English listener 300 will be discussed. If the talker 100 speaks in Korean "I'd like to confirm my reservation, please.", the simultaneous interpretation system 500 translates the wording into English and transmits an English voice, i.e. "I'd like to confirm my reservation, please." to the listener 300, corresponding to the translated wording. If the listener 300 replies "One moment, please.", the simultaneous interpretation system 500 translates the English reply of the listener 300 into Korean and transmits a Korean voice corresponding to the wording "One moment, please." to the talker 100.

In this embodiment of the present invention, it can be understood that the talker 100 and the listener 300 are users of communication terminals that can connect with the simultaneous interpretation system 500 through an IP network or the PSTN 700 such as a wired telephone, a mobile phone and a personal computer. In a case where the users connect with the simultaneous interpretation system 500 through a personal computer, a router (not shown) and a Voice over IP gateway (VoIP G/W) for connecting with the IP network (not

shown) connectable to the PSTN 700 may be further included in the users.

FIG. 4 is a view schematically showing the configuration of the third-party call control type simultaneous interpretation system according to the present invention. Referring to FIG. 4, the third-party call control type simultaneous interpretation system 500 of the present invention comprises a CTI board 510, a CTI control module 530, an interpretation module 550, and a main control module 570. The simultaneous interpretation system 500 is configured in such a manner that interactive simultaneous interpretation services can be provided to the talker 100 and listener 300 connected through the wired/wireless communication network by controlling the CTI control module 530 using the main control module 570.

Computer-Telephony Integration (CTI) is a technique for managing telephone calls using the computer. Main functions of the CTI include a voice store and forward function for recording and playing a voice input from a user, a digit capture function for recognizing dialing digits, and an out-dial function for dialing a specific telephone number to connect a call.

The CTI board 510 is configured to perform the above CTI functions, installed in the computer, and used to control a telephone circuit by connecting to the PBX. Since the CTI board 510 is identical to a CTI board commonly used in the automatic response system (ARS) in view of their configurations and operations, a detailed explanation thereof will be omitted.

The CTI control module 530 controls the CTI board 510 and the interpretation module 550 with the request of the main control module 570 and includes an event handler 531 for generating events in response to button signals input through the CTI board 510, a CTI application programming interface (API) 533 including CTI control functions for controlling the CTI board 510, and a working section 535 for calling the CTI control functions in order from the CTI API 533 with the request of the main control module 570 and performing basic telephone actions (e.g., dialing, answering and hanging up of the telephone).

The event handler 531 generates events in response to button signals input through the CTI board 510 and outputs messages according to the respective events to the main control module 570. For example, if it is detected that the telephone has been called from the talker 100 through the CTI board 510, the event handler 531 transmits an

EVT\_WAITCALL message to the main control module 570 according to the call reception.

The CTI API 533 is a telephony application program interface (TAPI) used for communication between the computer and the telephone, and can be understood as a kind of library in which the CTI control functions capable of controlling the CTI board 510 are stored.

5 When the CTI control functions are called, the CTI API 533 causes the CTI control functions to be decoded as command words comprehensible by the CTI board 510 and controls the CTI board 510 in accordance with the decoded command words. Here, TAPI available from Microsoft may be generally used as the CTI API.

10 Interfaces for the basic telephone actions such as out-dial, digit capture and voice recording can be provided through the CTI API 533. For example, when a telephone number of the listener 300 to which the talker 100 wishes to call is input, a DTMF tone detection function stored in the CTI API 533 is called so that the CTI API 533 can recognize the telephone number input by the talker 100.

15 The CTI control functions stored in the CTI API 533 will be more specifically explained as follows. The CTI control functions such as dx\_dial, dx\_sethook, dx\_getdig, dx\_fileopen, dx\_play and dx\_rec mean a dialing action, a hook setting action for answering or hanging up the phone, an action for detecting which buttons are pressed by the talker or listener, a file opening action, a file playing action, and a voice recording action, respectively.

20 However, since these CTI control functions are implemented to perform only a single function such as dialing, hook initialization, DTMF tone detection, and file playing, there is a disadvantage in that they should be separately and repeatedly called in order to perform the basic telephone actions such as the dialing, answering and hanging up of the telephone. Further, whenever the CTI control functions are called, the current state thereof should be confirmed and necessary CTI control functions should also be additionally requested.

25 For example, when the talker 100 inputs the telephone number of the listener 300, the simultaneous interpretation system 500 calls the CTI control function dx\_dial from the CTI API 533, generates a DTMF signal corresponding to the telephone number of the listener 300 through the CTI board 510, and attempts to connect the call. At this time, the CTI control functions to be executed later are determined according to whether the listener 300 can talk  
30 over the telephone. That is, if the tone signals are input from the telephone line of the

listener 300 through the CTI board 510, the simultaneous interpretation system recognizes that the talker 100 can talk over the telephone, and then, calls ATDX\_CPTerm as the following CTI control function and transmits ringing signals to the telephone of the listener 300. On the other hand, if a busy signal is input from the telephone line of the listener 300 through the CTI board 510, the simultaneous interpretation system recognizes that the listener 300 cannot talk over the telephone, and then, calls dx\_play as the following control function and outputs a call connection failure message. That is, in order to perform the phone dialing action, the CTI control function, dx\_dial, should be called and then the different CTI control functions should also be called in accordance with the signals input from the CTI board 510.

Therefore, in order to solve the above problems, the present invention is configured such that the CTI control functions are configured as a work unit capable of performing the basic telephone actions and are then called in order through the working section 535 to perform the basic telephone actions. Hereinafter, the working section 535 will be explained more in detail.

In general, a job means a unit of work that a computer can execute. In the present invention, the job can be understood as a sequence of CTI control functions configured to perform the basic telephone actions. An example of the basic telephone actions configured as a job unit is shown in FIG. 5.

Referring to FIG. 5, the jobs (JB\_\*) such as phone dialing, phone answering, phone disconnection or hanging up, button pressing, button reading, tone detection, voice forward, voice store, speaking and listening are configured as a sequence of CTI control functions. In particular, the CTI control functions in the shaded block are used to confirm the events generated from the event handler 531 or current state thereof and configured such that the following CTI control functions necessary at the next stage are called in response to the events generated from the event handler 531.

Therefore, since the CTI control functions are configured as a job unit as described above, the basic telephone actions can be made in accordance with only one job request without individually and repeatedly calling the CTI control functions. Accordingly, system control performance and speed can be improved.

In the meantime, the interpretation module 550 translates the voice of the talker 100

or listener 300 input from the CTI board 510 into a language recognizable by the other party, and includes a speech recognition section 551, a translation section 553, and a speech synthesis section 555.

The speech recognition section 551 recognizes the voice of the talker 100 or listener 300 input through the CTI board 510 and converts the recognized voice into a sentence (text). To this end, a hidden Markov model for calculating similarities between models using estimated values of the models obtained on the basis of changes in voice spectrums may be used as a speech recognition algorithm.

The translation section 553 translates the sentences recognized in the speech recognition section 551 into languages recognizable by the talker 100 or listener 300. To this end, the conventional rule-based translation algorithm through sentence analysis, lexical-based translation algorithm through language phenomenon, example-based translation algorithm through a large volume of examples, and the like can be used as they are. Thus, a detailed explanation thereof will be omitted.

The speech synthesis section 555 synthesizes the speech from the sentences which have been recognized from the speech recognition section 551 or translated from the translation section 553, and outputs the synthesized speech. To this end, a Holmant text-to-speech synthesis algorithm, which is disclosed in the technical paper "From Text to Speech" (Cambridge University Press, 1987, pp. 16-150) by J. Allen, M. S. Hunnicutt, D. Klatt et al., may be used as a text-to-speech algorithm.

Algorithms other than the aforementioned speech recognition algorithm, translation algorithm and text-to-speech synthesis algorithm may be used, and the present invention is not limited to these algorithms.

Furthermore, it cannot be known when any events will be generated from the talker 100 and the listener 300 in a kind of third-party call control type simultaneous interpretation system according to the present invention. Thus, in order to provide smooth interpretation services, actions necessary for the next stages should be able to be performed in accordance with the generated events.

To this end, the main control module 570 of the present invention controls the general operations related to the interactive simultaneous interpretation service based on an



interpretation scenario to be described later. Hereinafter, the main control module 570 will be explained more in detail.

The main control module 570 includes an interpretation scenario management section 571 for selecting the action to be executed in the next stage on the basis of a predetermined interpretation scenario when the events are generated in the CTI control module 530, and a state conversion section 573 for converting a current state into the next state in response to the current state conversion action selected from the interpretation scenario management section 571.

The interpretation scenario is an action flow of the simultaneous interpretation system 500, which has been beforehand defined such that a smooth simultaneous interpretation service can be provided to the talker 100 and the listener 300. The actions, which should be executed at the next stage in response to the events generated at the current state, are predetermined in the interpretation scenario of which one example is in turn illustrated in FIG. 6.

Referring to FIG. 6, the interpretation scenario is formulated in tables in the format of <'current state', 'event', 'action'>, wherein the 'current state' means an currently operating state (ST\_\*), the 'event' means a generated event (EVT\_\*), and the 'action' means an action (On\_\*) that should be performed at the next stage in response to the generated event. Further, the 'action' means an action for selecting the current state conversion action to convert the current state into the next state in response to the generated event and selecting the basic telephone actions necessary for the next stage.

That is, the interpretation scenario management section 571 selects the action (On\_\*) to be executed at the next stage on the basis of the previously stored interpretation scenario when events are generated from the event handler 531. If the interpretation scenario management section 571 selects an action (On\_\*), the current state conversion action and basic telephone action necessary for the next stage are selected in accordance with the selected action. Accordingly, the state conversion section 573 converts the current state into the next state in response to the selected current state conversion action, and the working section 535 performs the jobs necessary for the next stage in response to the selected basic telephone action.

For example, if the talker 100 connects with the simultaneous interpretation system 500, the event handler 531 of the CTI control module 530 transmits a call receiving event to the interpretation scenario management section 571 of the main control module 570. Then, the interpretation scenario management section 571 references <ST\_START, EVT\_WAITCALL, OnGotoPlayWelcomeMent> for processing the call receiving event from the interpretation scenario, converts the current state from ST\_START to ST\_PlayWelcomeMent by means of the state conversion section 573, and performs the action of outputting a connection welcoming message to the talker 100.

As mentioned above, since the interpretation scenario is configured in the format of <current state, event, action>, the action necessary for the next stage can be immediately performed regardless of what events are generated from the talker 100 and the listener 300 so that smooth communication between the talker 100 and the listener 300 who use different languages can be made.

Hereinafter, the third-party call control type simultaneous interpretation method of the present invention will be explained in detail with reference to the accompanying drawings.

FIG. 7 is a flowchart illustrating an entire process of the third-party call control type simultaneous interpretation method of the present invention, which comprises a telephone connection step (S10-S70) of establishing a traffic channel between the talker 100 and the listener 300 when the talker 100 connects with the simultaneous interpretation system 500, an automatic interpretation step (S80-S150) of translating the input voice of the talker 100 and the listener 300 into a language recognizable by the other party in accordance with a predetermined interpretation scenario, and an interpretation transmission step (S160-S170) of transmitting the translated voice of the talker 100 or the listener 300 to the other party in accordance with the interpretation scenario.

First, when the talker 100 calls a phone to connect with the simultaneous interpretation system 500, the call receiving event EVT\_WAITCALL is transmitted to the interpretation scenario management section 571 through the event handler 531. At this time, the interpretation scenario management section 571 selects the action OnGotoPlayWelcomeMent for processing the call receiving event in accordance with <ST\_START, EVT\_WAITCALL, OnGotoPlayWelcomeMent> of the interpretation scenario,

converts the current state into a welcome message output state by means of the state conversion section 573 according to the selected action OnGotoPlayWelcomeMent, and performs the phone answering action by means of the working section 535 (S10). Here, since the operations of the event handler 531, the working section 535, the interpretation scenario management section 571, and the state conversion section 573 have been explained in detail in connection with FIG. 4, they will be briefly described together with the simultaneous interpretation system 500 for the convenience of explanation.

Next, after the phone answering action has been completed, the simultaneous interpretation system 500 outputs a welcome message in accordance with <ST\_PLAYWELCOMEMENT, EVT\_PLAYVOICE, OnEndPlayWelcomeMent> of the interpretation scenario (S20). Then, the system outputs a message requesting the input of the telephone number of the listener 300 in accordance with <ST\_PLAYPHONENUMMENT, EVT\_PLAYVOICE, OnEndPlayPhoneNumMent> of the interpretation scenario (S30).

When the talker 100 inputs the digits through the telephone, the DTMF tone signal event EVT\_GETDIGIT is produced. Thus, the simultaneous interpretation system 500 detects the DTMF tone signals input from the talker 100 and recognizes the telephone number of the listener 300 in accordance with <ST\_GETPHONENUMDIGIT, EVT\_GETDIGIT, OnEndGetPhoneNumDigit> of the interpretation scenario (S40).

After the telephone number of the listener 300 has been recognized as such, the simultaneous interpretation system 500 outputs the call connection announcement to the talker 100 and simultaneously performs the phone dialing action to attempt to connect the call to the telephone number of the listener 300 in accordance with <ST\_PLAYOUTBOUNDCALLMENT, EVT\_PLAYVOICE, OnEndPlayOutboundCallMent> of the interpretation scenario (S50).

Then, the interpretation system 500 determines whether the call has been connected based on whether the listener 300 has replied to the call. If the call connection has failed, the interpretation system 500 outputs the call connection fail message to the talker 100 in accordance with <ST\_PLAYCONNECTFAILEMENT, EVT\_PLAYVOICE, OnEndPlayConnectFailMent> of the interpretation scenario (S60). On the other hand, if the call connection has succeeded, the interpretation system outputs the call connection success

message to the talker in accordance with <ST\_PLAYCONNECTSUCESSMENT, EVT\_PLAYVOICE, OnEndPlayConnectSucessMent> of the interpretation scenario (S70).

In a case where the call connection has succeeded, i.e., the call receiving event has been generated, the simultaneous interpretation system 500 outputs a use announcement for use in the interpretation services to the talker 100 and the listener 300 in accordance with <ST\_PLAYINTRODUCTION, EVT\_PLAYVOICE, OnEndPlayIntroduceMent> (S80).

In the meantime, the simultaneous interpretation system 500 according to the present invention controls two traffic channels between the talker 100 and the simultaneous interpretation system 500 and between the simultaneous interpretation system 500 and the listener 300 at the same time so that the interpretation services can be provided in real time to both the talker 100 and the listener 300. Since the interpretation system of the present invention controls these two traffic channels at the same time according to the same interpretation scenario, only a case where the traffic channel between the talker 100 and the simultaneous interpretation system 500 is controlled will be described by way of example for the convenience of explanation.

After the use announcement for use in the interpretation services has been output, the simultaneous interpretation system 500 records the voice input by the talker 100 in accordance with <ST\_GETRECOGSTARTDIGIT, EVT\_PLAYVOICE, OnEndGetRecogStartDigit> of the interpretation scenario when the talker 100 presses a predetermined button (e.g., \* button) for his/her speech input (S90).

When the talker 100 presses a predetermined button (e.g., # button) to terminate a recording process during the voice recording, the simultaneous interpretation system 500 terminates the recording of the voice of the talker 100 in accordance with <ST\_GETRECOGSTOPDIGIT, EVT\_PLAYVOICE, OnEndGetRecogStopDigit> of the interpretation scenario (S100).

Then, the simultaneous interpretation system 500 recognizes the recorded voice or speech of the talker 100 in accordance with <ST\_SPEECHRECOG, EVT\_RECOGSPEECH, OnEndSpeechRecog> of the interpretation scenario (S110). As a result, if speech recognition has failed, the simultaneous interpretation system outputs the speech recognition fail message in accordance with <ST\_PLAYRECOGFAILMENT, EVT\_PLAYVOICE,

OnEndPlayRecogFailMent> of the interpretation scenario and then returns to a state where it is ready to receive the voice of the talker 100 (S120). If the speech recognition has succeeded, the system synthesizes the speech from the recognized sentence in accordance with <ST\_PLAYTTSRECOGSENTENCE, EVT\_PLAYVOICE, OnEndPlayTtsRecogSentence> and then transmits the speech to the talker 100 (S130).

When the recognized sentence synthesized into speech is transmitted to the talker 100, the talker 100 confirms whether his/her input contents are correct. The talker 100 selects the \* button if the input contents are correct, whereas the talker selects the \* button if the contents are incorrect. In a case where the talker selects the \* button, the simultaneous interpretation system 500 translates the recognized sentence into a language recognizable by the listener 300 in accordance with <ST\_TRANSRECOGSENTENCE, EVT\_TRANS, OnEndTransRecogSentence> of the interpretation scenario (S140). After the translation has been completed, the interpretation system 500 synthesizes the translated sentence into the speech and outputs the speech to the listener 300 in accordance with <ST\_PLAYTTSTRANSSENTENCE, EVT\_PLAYVOICE, OnEndPlayTtsTransSentence> of the interpretation scenario (S150).

Next, the simultaneous interpretation system 500 transmits the translated voice of the talker 100 to the listener 300 in accordance with <ST\_OUTTRANSSENTENCE, EVT\_PLAYVOICE, OnEndOutTransSentence> of the interpretation scenario (S160). After the synthesized speech of the translated sentence has been output, a predetermined alarm sound (e.g., dingdong) indicative of the termination of sound output may be output in accordance with <ST\_PLAYDINGDONGMENT, EVT\_PLAYVOICE, OnEndPlayDingdongMent> of the interpretation scenario.

Next, the simultaneous interpretation system 500 checks whether there is a reply to the transmitted voice from the listener 300 in accordance with <ST\_PLAYRCVWAITMENT, EVT\_RCVSENTENCE, OnEndGetRcvSentence> of the interpretation scenario (S170). If an answer sentence is received from the listener 300, the simultaneous interpretation system 500 transmits the answer sentence to the talker 100 in accordance with <ST\_OUTRCVSENTENCE, EVT\_PLAYVOICE, OnEndOutRcvSentence> of the interpretation scenario (S180).

As described above, the simultaneous interpretation system 500 of the present invention controls all the operations associated with the interactive simultaneous interpretation services in accordance with the interpretation scenario in which the actions to be performed at the next stages are defined beforehand. Therefore, the talker 100 can freely speak by telephone with the listener 300 who uses a different language and is remotely located.

According to the third-party call control type simultaneous interpretation system and method of the present invention, communication between different language users can be smoothly made without purchasing additional specific terminals. Thus, there is an advantage in that the simultaneous interpretation services can be used at a low cost.

Although the present invention has been described in connection with the preferred embodiments shown in the drawings, it will be apparent to those skilled in the art that various changes and modifications can be made thereto without departing from the scope and spirit of the present invention. Therefore, the true scope of the present invention should be defined by the appended claims.